FEBS Letters 582 (2008) 1169–1169

Preface

Finally: The digital, democratic age of scientific abstracts

Welcome to the first issue of FEBS Letters with structured digital abstracts. A common complaint among researchers in the postgenomic era of orgiastic data production is our inability to cope with the scientific output. Much more data are published in the primary literature than any of us can digest. It is estimated that a large proportion of publicly funded research is on problems and objects that have been published or patented before; people are simply not aware of previous work due to ambiguous designations. More often than not, molecules have not only two, but multiple names. We all have read beautiful papers where, while if it may be clear what protein or gene it is that is the object of the study, it is impossible to know, for example, what species it is that is being investigated, or the exact cell type.

Every year, even every day, that passes without the biological information that is obtained and published by tens of thousands of laboratories worldwide being machine-readable is a lost opportunity and a costly affair. Essentially, if there are no digital standards on how information can be delivered, the information is not really accessible in the sea of published papers.

Another important aspect is that scientists from different backgrounds and nations report on the same scientific content with various degrees of literary precision and packaging skills. As a result, there is a strong cultural bias to the efficacy and impact of scientific reporting. The unfairness of this is obvious to the community but few measures have been taken to offset this. Ideally, the process of creating and retrieving scientific data should be made equally accessible to everybody so that historical cultural advantages wane with time to give way to a more democratic forum and procedure. Importantly, it should be the author and not the database curator that decides what is the essence of the published paper.

Hence this initiative at FEBS Letters. FEBS Letters has a broad base of contributors and readers across the world and publishes short topical papers usually on a single type of molecular observation. It is therefore an ideal journal to start structuring the abstracts of its scientific papers in a way that is machine-readable and not ambiguous.

Obviously, it is impossible to jump start the digitalization of complex and diverse scientific content, so we decided to begin there where parameters were sufficiently clear to allow for an experiment: the reporting of protein interactions and post-translational modifications. In 6 months, this "experiment" will be evaluated. If it is successful, eventually all appropriate FEBS Letters manuscripts will obtain an structured digital abstract (SDA).

With this issue, we inaugurate the first five papers that contain a SDA. We congratulate the authors as they have started what we believe is the beginning of a revolution.

It is our conviction that the ability of publishing SDAs should strongly raise the attractiveness of publishing in FEBS Letters because for the first time the authors will know that their paper can be "read", i.e. "understood" by machines, computers, and therefore can be analysed, stored and cited orders of magnitudes better and more precisely than has been possible so far. To accompany this first issue, we have asked for contributions of like-minded colleagues that have been advocating the creation of SDA for some time, have worked on the a *posteriori* extraction and digitalization of data operated through text-mining procedures, and have curated databases and web services to assist the community in the management of data. Through their contribution, we hope that the reader obtains the necessary contextualization of the challenges and the opportunities associated with the transition to the digital era. Clearly, for years to come, text-mining is likely to represent the main avenue of extraction of digital data from the scientific literature. At the same time, the creation of future, more comprehensive digital abstracts will have to be guided by the experience obtained from text-mining and natural language studies. It is going to be interesting to monitor the parallel developments, cross-fertilizations, and benchmarking. Ceol et al. provide more detail and background behind this SDA experiment and illustrate the relationship with MINT and other protein interaction databases. Seringhaus and Gerstein, who are among the early advocates for this process, review the pros and cons of having the authors involved in the process to different degrees and discuss when and how text-mining should be applied. Lastly, Leitner and Valencia assess the current performance of text-mining tools and their potential for creating a SDA in the first place.

This initiative and issue would not have been possible without a visionary society (FEBS) and publisher (Elsevier) that has accepted the risk of a slow or even false start in the digitalization of scientific information. We all should be grateful for this opportunity. Now it is up to our community to vote for this transformation by adopting it, improving it and exporting it to other journals. We all know there is no way back to the dark pre-digital ages. Let there be light!

Giulio Superti-Furga
CeMM Center for Molecular Medicine of the Austrian Academy of Sciences Vienna, Austria
E-mail address: gsuperti@cemm.oeaw.ac.at

Felix Wieland
Heidelberg University Biochemistry Center Heidelberg, Germany
E-mail address: felix.wieland@bzh.uni-heildelberg.de

Giovanni Cesareni
University "Tor Vergata", Rome, Italy
E-mail address: cesareni@uniroma2.it

doi:10.1016/j.febslet.2008.02.070